Hierarchical Bayesian Selection Procedures for the Best Binomial Population*

by

J.J. Deely
University of Canterbury

S.S. Gupta
Purdue University

Technical Report # 88-21C

PURDUE UNIVERSITY

CENTER FOR STATISTICAL DECISION SCIENCES AND DEPARTMENT OF STATISTICS

(1)

Hierarchical Bayesian Selection Procedures
for the Best Binomial Population*

by

J.J. Deely
University of Canterbury

S.S. Gupta
Purdue University

Technical Report # 88-21C

Department of Statistics
Purdue University

May, 1988

* This research was supported in part by the Office of Naval Research Contract N00014-88-K-0170 and NSF Grants DMS-8606964 and DMS-8702620 at Purdue University.

HIERARCHICAL BAYESIAN SELECTION PROCEDURES
FOR THE BEST BINOMIAL POPULATION

by

J.J. Deely
University of Canterbury
and
S.S. Gupta
Purdue University

(to the k p

Abstract

In this paper a hierarchical Bayesian ~~(HB)~~ model is adopted to derive selection procedures for selecting the best of k binomial parameters, say the probability of success corresponding to k different suppliers. This model facilitates the use of prior information in the analysis for both small and large sample sizes. In addition to computing posterior probabilities that the i^{th} supplier is best, this paper presents expressions for deciding how much better a given supplier is relative to the others. Prior information is assumed to begin with exchangeability and can be more informative if the experimenter has other knowledge about the suppliers as a group. A numerical example is given and the paper concludes with remarks about future work.

AMS 1980 Subject Classification: 62F07, 62C12.

Key words and phrases: binomial, Bayes, hierarchical, selection procedures, exchangeable, hyperpriors, empirical Bayes

Hierarchical Bayesian Selection Procedures
for the Best Binomial Population

1. Introduction

Suppose there are k suppliers of a particular item and a sample of n_i items is taken from the i^{th} supplier yielding x_i the number of successes (or defectives) in the sample. Then X_i has a binomial distribution with parameter θ_i, which denotes the true unkown probability of a success (failure) from the i^{th} supplier. The *best* supplier is defined to be the one with largest (smallest) θ_i. Based on the observed data and prior information available, we seek procedures which will select a non-empty subset of the k suppliers and assert with some confidence that the best supplier is amongst those in the selected subset. In this paper a hierarchical Bayesian (HB) model is adopted and the behaviour of various selection procedures thus obtained is studied. This application to binomial data parallels the normal means problem considered by Berger and Deely (1988).

The problem of selecting the best binomial population has received considerable attention in the literature mainly from a non-Bayesian approach. Pioneering papers by Sobel and Huyett (1957) and Gupta and Sobel (1960) dealt with selecting the best and selecting a subset containing the best binomial population respectively. Later, Gupta, Huang and Huang (1976) studied a conditional subset selection rule and a related test of homogeneity. A good discussion of these and other non-Bayesian papers can be found in books by Gibbons, Olkin and Sobel (1977) and Gupta and Panchapakesan (1979). It is not the purpose of this paper to discuss the relative merits of the non-Bayesian vs Bayesian approaches, but we believe that for the binomial selection problem studied in this paper, the Bayesian model contains the facility to deal easily with the type of information which is likely to occur in practice and in that sense offers the practitioner a more appealing model.

There have been some Bayesian and empirical Bayesian papers dealing with the binomial selection problem. Deely (1965) developed empirical Bayes procedures for general selection problems including among them the binomial case with independent θ_i's, each with a beta prior with unknown parameters. Gupta and Liang (1986) derived non-parametric empirical Bayes procedures for selecting the best binomial population under the assumption that $\theta_1, \ldots, \theta_k$ are independent each with an unknown non-parametric prior distribution. Bratcher and Bland (1975) considered a naive Bayesian approach in which the θ_i's are independent with known but perhaps different beta priors. They considered various multiple comparisons based on computing the posterior probabilities of each population being best and used numerical integration to calculate these. Later Yang (1987) applied their model but adopted the so called PP^* criterion, which had been previously introduced for a general selection problem by Gupta and Yang (1985). This criterion, in an effort to relate the Bayesian criterion to the classical P^* condition, states that the Bayes P^* procedure selects the smallest subset for which the posterior probability that the subset contains the best is at least P^*.

There have been other relevant papers dealing with estimation as opposed to selection for the binomial case. Albert (1984) considers the simultaneous estimation of k binomial probabilities and develops empirical Bayes estimators under an exchangeable hierarchical model. Leonard (1972) also considers this problem but uses a logit transformation to bring the problem into a multivariate normal context. A lot acceptance problem was considered by Eaves (1980) in which n items are drawn from each of k lots under binomial sampling. An exchangeable hierarchical model is assumed and the predictive distribution for the next lot is computed when all items from all lots are good.

A related problem, that of allocating the observations to the various suppliers con-

strained by the fact that the total is fixed, has received some attention in the literature. Brooks (1987) deals with a Bayesian approach for $k = 2$, while Brittain and Schlesselman (1982) discuss this case from a frequentist viewpoint when trying to estimate $p_1 - p_2$ or p_1/p_2. These problems will not be discussed but some conclusions about allocation can be drawn from the work presented here and these will be discussed in Section 5.

The approach in this paper is to present a model which has the capability of incorporating prior information concerning the suppliers as a group into the analysis. The literature to date while recognizing the usefulness of such prior information in other problems (see for example Berger (1985), Chapter 3) has ignored applying such models to the binomial selection problem. The hierarchical Bayesian model is one way this can be done, easily and with useful results. These ideas are discussed more thoroughly in Section 3 after having presented the mathematical details of the model in Section 2. An example illustrating various aspects of the model is given in Section 4 with concluding remarks and suggestions for further work given in Section 5.

2. Mathematical details, the prior distribution and selection criteria

Let $\underline{x} = (x_1, \ldots, x_k)$ be the vector of observations from the k suppliers, x_i conditional on θ_i having the binomial distribution

$$f(x_i|\theta_i) = \binom{n_i}{x_i} \theta_i^{x_i}(1-\theta_i)^{n_i - x_i},$$

and let $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ be the vector of unknown parameters for which we want to select that supplier with largest θ_i. The prior distribution $\pi(\underline{\theta})$ on $\underline{\theta}$ will be obtained via the hierarchical Bayesian structure (see Berger (1985), Section 4.6) in which $\pi(\underline{\theta})$ is given as a mixture of a prior conditional on hyperparameters β and η with a hyperprior

distribution on these parameters; that is,

$$\pi(\underline{\theta}) = \int\int \pi(\underline{\theta}|\beta,\eta)h(\beta,\eta)d\beta d\eta.$$

Conditional upon the hyperparameters β, η, the components of $\underline{\theta}$ are assumed to be i.i.d. with a common beta distribution given by

$$\pi(\theta_i|\beta,\eta) = \frac{\Gamma(1/\eta)}{\Gamma(\beta/\eta)\Gamma((1-\beta)/\eta)}\theta_i^{\frac{\beta}{\eta}-1}(1-\theta_i)^{\frac{1-\beta}{\eta}-1} \tag{2.1}$$

where $0 < \beta < 1, \eta > 0$; thus

$$\pi(\underline{\theta}|\beta,\eta) = \prod_{i=1}^{k} \pi(\theta_i|\beta,\eta).$$

This partricular form of the beta distribution will be convenient for the numerical computations and elicitation of prior information. These topics will be discussed more fully in the next section. Special note is taken here that

$$E(\theta_i|\beta,\eta) = \beta \text{ and } \sigma^2 = \text{ Var } (\theta_i|\beta,\eta) = (1-\beta)(\eta/(\eta+1)) \tag{2.2}$$

The hyperparameters β and η willbe taken as independent so that

$$h(\beta,\eta) = h_1(\beta)h_2(\eta)$$

in which h_1 will be some member of the beta family and h_2 as some member of the family given by

$$h_2(\eta) = \begin{cases} \frac{m-1}{4\rho mc}(\eta+1)^{-2}, & 0 < \eta < 4c/(1-4c) \\ \frac{m-1}{4\rho mc}(\eta+1)^{m-2}(4c/\eta)^m, & \eta \geq 4c/(1-4c) \end{cases} \tag{2.3}$$

where ρ is the normalizing constant, c and m are parameters, $0 < c < 1/4$ and $m \geq 2$. The values of the parameters for h_1 and h_2 will depend upon what prior information is

available, the noninformative exchangeable case being h_1 uniform and h_2 having $c = 1/4$ and $m = \infty$.

We will use the following notation for the beta distribution:

$$g(y|a,b) = B(a,b)y^{a-1}(1-y)^{b-1}$$

$$B(a,b) = \Gamma(a+b)/(\Gamma(a)\Gamma(b))$$

$$G(t|a,b) = \int_0^t g(y|a,b)dy.$$

Under the notation and assumptions above, it follows that the conditional distribution of $\underline{\theta}$ given $\underline{x}, \beta$ and η is given by

$$\pi(\underline{\theta}|\underline{x},\beta,\eta) = \prod_{i=1}^{k} \pi(\theta_i|x_i,\beta,\eta) \tag{2.4}$$

where

$$\pi(\theta_i|x_i,\beta,\eta) = \frac{f(x_i|\theta_i)\pi(\theta_i|\beta,n)}{f(x_i|\beta,\eta)} = g(\theta_i|a_i,b_i),$$

$$f(x_i|\beta,\eta) = \int_0^1 f(x_i|\theta_i)\pi(\theta_i|\beta,\eta)d\theta_i = \binom{n_i}{x_i} B(a,b)/B(a_i,b_i)$$

and $a = \beta/\eta, b = (1-\beta)/\eta, a_i = a + x_i, b_i = b + \eta_i - x_i$. Let

$$f(\underline{x}|\beta,\eta) = \prod_{i=1}^{k} f(x_i|\beta,\eta) \text{ and } f(\underline{x}) = \int_0^\infty \int_0^1 f(\underline{x}|\beta,\eta)h_1(\beta)h_2(\eta)d\beta d\eta. \tag{2.5}$$

Then the posterior distribution of $\underline{\theta}$ given the data $\underline{x}$ can be written as

$$\pi(\underline{\theta}|\underline{x}) = \int_0^\infty \int_0^1 \pi(\underline{\theta}|\underline{x},\beta,\eta)\frac{f(\underline{x}|\beta,\eta)}{f(\underline{x})}h_1(\beta)h_2(\eta)d\beta d\eta. \tag{2.6}$$

In fact we will not require the precise form of $\pi(\underline{\theta}|\underline{x})$ since decisions about which supplier or subset of suppliers should be selected will be based on easily computed expectations taken with respect to this posterior. We now develop two such criteria.

(i) Posterior probability of getting the best.

Let

$$p_j(b) = P(\theta_j > \theta_i + b \text{ for all } i \neq j|\underline{x}),$$

where $b \in [0,1]$. It will be noted that $p_j(0)$ is just the posterior probability that θ_j is largest and the usual PP^* selection criterion of Gupta and Yang is obtained by putting in the selected subset the smallest number of suppliers for which the sum of their corresponding $p_j(0)$'s is at least P^*. We have here suggested a stronger criterion for selection purposes; one that allows the practitioner to express a quantity b, i.e. how superior does the best have to be, and a probability P^* to be attained by the selected subset. Of course for $b > 0$ it is no longer true that $\Sigma p_j(b) = 1$ and in fact it may be that for a given $b > 0$ no supplier is better than the others by amount b with high enough probability. The experimenter can easily take another look and perhaps lower b or the probability requirement. In any case we believe the $p_j(b)$'s provide a useful criterion for selecting one or more suppliers and gives the experimenter the interpretation which relates to the practical problem.

Using (2.1) and letting $A_j(b) = \{\underline{\theta} : \theta_j > \theta_i + b \text{ for all } i \neq j\}$, the expression for $p_j(b)$ can be derived as follows:

$$\begin{aligned} p_j(b) &= \int_{A_j(b)} \pi(\underline{\theta}|\underline{x})d\theta \\ &= \int_0^\infty \int_0^1 \left[\int_{A_j(b)} \pi(\underline{\theta}|\underline{x},\beta,\eta)d\underline{\theta} \right] \frac{f(\underline{x}|\beta,\eta)}{f(\underline{x})} h_1(\beta)h_2(\eta)d\beta d\eta \end{aligned}$$

$$(2.7) \qquad = \int_0^\infty \int_0^1 \left[\int_0^1 \prod_{\substack{i=1 \\ i\neq j}}^{k} G(\theta_j - b|a_i,b_i)g(\theta_j|a_j,b_j)d\theta_j \right] \frac{f(\underline{x}|\beta,\eta)}{f(\underline{x})} h_1(\beta)h_2(\eta)d\beta d\eta$$

noting that the terms in brackets are equal, a_i and b_i being defined earlier. Thus evaluation of each $p_j(b)$ requires only a three dimensional numerical integration for all choices of h_1 and h_2, provided the incomplete beta function is available.

(ii) Expected number of future successes.

Another useful criterion for selection purposes is obtained by considering a future observation. Suppose n future observations are to be taken from one of the suppliers. Let the total number of successes be denoted by Y and compute $E(Y_i)$ for each supplier $i = 1, \ldots, k$, where E is the expection taken with respect to the distribution of Y conditional on $\underline{x}$, i.e. the predictive distribution. Then the supplier with largest $E(Y_i)$ is called best. Calculation for $E(Y_i)$ is easily obtained as

$$\begin{aligned} E(Y_i) = E(Y_i|\underline{x}) &= \int_0^1 E(Y_i|\underline{x}, \theta_i)\pi(\theta_i|\underline{x})d\theta_i \\ &= \int_0^1 n\theta_i\pi(\theta_i|\underline{x})d\theta_i \\ &= nE(\theta_i|\underline{x}). \end{aligned}$$

Thus ranking suppliers on the basis of largest expectation of the number of successes in n future items is equivalent to ranking them on the basis of their posterior means based on the present data $\underline{x}$. An expression for $E(\theta_i|\underline{x})$ involves only a two dimensional integration and is given by:

$$\begin{aligned} (2.8) \qquad E(\theta_i|\underline{x}) &= \int_0^1 \theta_i\pi(\theta_i|\underline{x})d\theta_i \\ &= \int_0^\infty \int_0^1 \left[\int_0^1 \theta_i\pi(\theta_i|x_i, \beta, \eta)d\theta_i\right] \frac{f(\underline{x}|\beta, \eta)}{f(\underline{x})} h_1(\beta)h_2(\eta)d\beta d\eta \\ &= \int_0^\infty \int_0^1 \left[\frac{\beta}{1+\eta n_i} + \frac{\eta n_i}{1+\eta n_i}\left(\frac{x_i}{n_i}\right)\right] \frac{f(\underline{x}|\beta, \eta)}{f(\underline{x})} h_1(\beta)h_2(\eta)d\beta d\eta \end{aligned}$$

using (2.4), (2.5) and noting that the mean of $g(\theta_i|a_i, b_i)$ is $a_i/(a_i + b_i)$.

Using the posterior means to rank the suppliers, an appropriate decision about which subset of suppliers to select can be made, e.g. put in the selected subset supplier i if and only if $E(\theta_i|\underline{x}) \geq c$. This selection procedure assures the decision maker that each of the suppliers thus selected will have the expected number of successes at least as large as nc.

Alternatively, if one selects a subset of size $r, r < k$, based on the r largest $E(\theta_j|\underline{x})$, then the expected number of successes for that subset is larger than that for any other subset of size r. Further amplification of this point will be made in Section 5. We now turn our attention to the various choices for h_1 and h_2 and discuss the influence of prior information on these choices.

3. Prior information and elicitation for h_1 and h_2

There are two main advantages of the seemingly complicated hierarchical structure. Firstly, it provides a realistic Bayesian model which can easily accommodate the type of prior information which is likely to be available; secondly, it is the appropriate model for what is commonly called the parametric empirical Bayes approach (see Morris (1983)). In the particular application made herein to supplier's data, it is clear that there is some prior information concerning the suppliers as a group, i.e. approximately where their quality is likely to be and what sort of variability amongst the θ_i's can be expected. But if this kind of information is unavailable, then it is still sensible to treat the θ_i's as exchangeable with noninformative hyperpriors. Both of these ideas are covered in the HB model. This type of prior information is to be contrasted to those Bayesian models which assume the θ_i's are independent with known but perhaps different distributions. This approach is generally quite unrealistic and therefore has limited application. On the other hand it is sometimes argued that a prior distribution on the θ_i's exists but is *unknown*. When this prior is assumed to be in some parametric family, it is then suggested that repetitions of the process may yield estimates of these parameters. Acting as though these estimates were the *true* unknown parameter values, one can then use the Bayes procedures, hence the expression parametric empirical Bayes. What estimators are sensible in this context is generally answered by embedding the unknowns in a larger truly Bayesian model, hence

the incorporation of hyperpriors and the expression Bayes empirical Bayes, (see Deely and Lindley (1981)). Often such truly Bayesian models yield complicated numerical problems related to the form of the posterior distribution, and in this case some form of Bayesian estimation is required. Again the HB model provides a structure within which sensible estimates are easily obtained, but we point out that the particular problem treated in this paper, no such estimators are required since $p_j(b)$ in (2.7) and $E(\theta_i|\underline{x})$ in (2.8) are easily computed.

We now discuss the choices for h_1 and h_2 and their relationship to the form of the prior information available.

<u>Case 1</u> : Exchangeable and noninformative.

In the situation in which practically nothing is known a priori about the suppliers with respect to $\underline{\theta}$, it is reasonable to assume that $\theta_1, \ldots, \theta_k$ are exchangeable random variables. Any prior distribution obtained via a mixture implies exchangeability and so in particular the structure given in the previous section insures exchangeability for any choice of h_1 and h_2. Consistent with the absence of prior information is the assumption of noninformative hyperpriors. Since $0 < \beta < 1$ we can take a noninformative choice for h_1 as the uniform distribution. For h_2 we argue that since $0 <$ Var $(\theta_i|\beta, \eta) < \frac{1}{4}$ for all β we deduce a noninformative choice for h_2 by putting a uniform distribution on the variance over $(0, \frac{1}{4})$. This gives

$$h_2(\eta) = \frac{1}{(\eta+1)^2}, 0 < \eta < \infty.$$

It could be suggested that a simple noninformative choice for h_2 would be $h_2(\eta) \equiv 1$. This has been used in the example in the next section for comparison purposes but in the special case in which the n^{th} component of the data vector $\underline{x}$ is either 0 or n_i for all $i = 1, 2, \ldots, k$,

the improper $h_2(\eta) \equiv 1$ does not yield a proper posterior. Also since the elicited prior information will concern the conditional mean and variance it will be more convenient to think of hyperpriors induced via this information or lack of it.

Case 2 : Prior information available.

It could be the case that some decision makers may have enough prior information to specify precise values for β and η in $\pi(\theta|\beta,\eta)$, that is, select a particular beta distribution as a prior distribution for $\theta_1,\dots,\theta_k$. In fact some parametric empirical Bayes models assume each θ_i is independently generated from a particular prior with unknown parameters β_i and η_i. However it has been recognized that this is a rather naive view of Bayesian models and that the notion of exchangeability amongst the θ_i's is a more realistic approach, (see, for example, Berger (1985), Chapter 4). Our approach here is to consider prior information arising from eliciting answers from the practitioner to the following questions:

(1) Where do you expect the average of the θ_i's to lie, i.e. can you specify an interval, say (s_1,t_1), within which you are confident that the average of the θ_i's will lie?

(2) How variable do you consider the θ_i's to be; that is, can you specify an interval, say (s_2,t_2), within which you are confident *all* of the θ_i's must lie?

Answering the first question will determine $h_1(\beta)$ as a member of the beta distribution whose mean is taken as the midpoint of the interval (s_1,t_1) and variance as $[(t_1-s_1)/4]^2$. This choice is influenced by convenience but it is consistent with the elicited information while also allowing a small probability that the mean of the θ_i's is outside the interval specified by the experimenter. Computation of $h_1(\beta|\underline{x})$ could be used to assess the experimenters original judgment.

We will use the answer to the second question to determine $h_2(\eta)$ by firstly using

this information to obtain an appropriate distribution on σ^2, the conditional variance of θ_i given β and η. Since the elicited information expresses an upper bound, say c, on σ^2 over all β and η, we take this to imply a flat distribution on the interval $(0, c)$. However we allow the possibility that the variance could exceed this value but with a distribution that decays exponentially to 1/4. Note that it is always the case that $0 < \sigma^2 < 1/4$. This distribution is called the 'shoe' distribution and is given by

$$s(u) = \begin{cases} \frac{m-1}{\rho m c}, & 0 < u < c \\ \frac{m-1}{\rho m c}\left(\frac{c}{u}\right)^m, & c \leq u < \frac{1}{4} \end{cases}$$

where $\rho = 1 - (4c)^{m-1}/m$ is the normalizing constant, c will be taken as $c = [(t_2 - s_2)/4]^2$ and m is chosen so that the $P(0 < \sigma^2 < c)$ describes the confidence of the practitioner. Observe that $P(0 < \sigma^2 < c) = \frac{m-1}{\rho m}$, so the determination of m is straightforward. From this distribution on σ^2 and using the transformation $\sigma^2 = \cdot/4(\eta + 1)$ it is easy to obtain h_2 as given in (2.3).

These hyperpriors will be used in a numerical example in Section 4 to show the effect on the selection criteria. We remark that other hyperpriors satisfying the information elicited were used but did not have much effect on $p_j(b)$ or $E(\theta_j|\underline{x})$.

Numerical Examples

In this section we study the effect of the hyperpriors and give examples of the relevant computations.

(i) Effect of sample size on $p_j(b)$

The table below compares the values of $p_j(b)$ when the sample size changes from 10 to 20 for both the noninformative and informative cases.

Table 1 – Values of $p_j(b)$; top figures for $n_i = 10$; figures in ( ) for $n_i = 20$.

		Noninformative			Informative		
Supplier	$\frac{x_i}{n_i}$	b			b		
		0	.05	.10	0	.05	.10
1	.1	.006 (.000)	.002 (.000)	.001 (.000)	.014 (.001)	.006 (.000)	.002 (.000)
2	.4	.194 (.110)	.136 (.060)	.088 (.029)	.217 (.120)	.144 (.063)	.088 (.030)
3	.6	.800 (.894)	.720 (.817)	.628 (.713)	.768 (.868)	.672 (.777)	.564 (.667)

These figures show that as the sample size increases, $p_j(b)$ for the largest sample proportion increases but the difference from noninformative to informative is not very large in either case. The noninformative hyperpriors were as in Case 1 of Section 3. For the informative case, hypothetical answers to Questions 1 and 2 as discussed in Case 2 of Section 3 were taken respectively as:

(1) $(s_1, t_1) = (0.3, 0.5)$ and thus h_1 was taken as $g(\beta|38, 57)$ where the parameters were taken as the solutions to

$$\frac{(.3+.5)}{2} = \frac{a}{a+b}; \left(\frac{.5-.3}{4}\right)^2 = \frac{ab}{(a+b)^2(a+b+1)};$$

(2) $(s_2, t_2) = (0.1, 0.6)$ and thus h_2 is given by (2.3) with $c = \frac{.6-.1}{4} = .125$ and $m = 4$.

Other choices for these informative hyperpriors were made with very little effect on the $p_j(b)$.

(ii) All sample proportions equal but unequal sample sizes

A useful feature of the hierarchical model is its ability to deal with equal proportions from the k suppliers when the sample sizes are in fact different. It seems to be the case that the smallest sample size always has the largest $p_j(0)$. The table below shows the computations for $p_j(0)$ when h_1 and h_2 are the noninformative choices. Other hyperpriors

did not result in much change in $p_j(0)$.

Table 2 – Values of $p_j(0)$ and $E(\theta_j|\underline{x})$; Equal proportions but unequal sample sizes.

j	n_j	$\frac{x_i}{n_i} = .2$	$\frac{x_i}{n_i} = .4$	n_i	$\frac{x_i}{n_i} = .45$	n_i	$\frac{x_i}{n_i} = .45$
1	10	.384	.375 (.407)	20	.374	40	.351 (.45)
2	20	.323	.325 (.405)	40	.325	60	.335 (.45)
3	30	.296	.302 (.404)	60	.302	80	.326 (.45)

The situation is quite different when using the posterior mean as the selection criterion. The numbers in parantheses are the corresponding posterior means and clearly provide no discrimination. This is to be expected since $E(\theta_j|\underline{x})$ in (2.8) is seen to be a convex combination of the sample proportion and the posterior mean of β. But β is centered around the average of the proportions; hence the convex combination of the two will be very nearly the one value of sample proportions. Again very little effect was obtained by using informative hyperpriors.

(iii) Unequal sample sizes and unequal proportions

The striking advantage of the hierarchical model is best displayed when dealing with variable sample sizes and unequal proportions. The table below indicates the type of computations possible for this model.

Table 3 – Values of $p_j(b)$ and $E(\theta_j|\underline{x}); h_1, h_2$ noninformative.

j		1	2	3	4	5	6	7	8
n_j		18	19	21	23	16	20	22	17
x_j		3	3	3	3	2	2	2	1
x_j/n_j		.167	.156	.143	.130	.125	.100	.091	.059
$E(\theta_j|\underline{x})$		.175	.167	.153	.141	.140	.116	.107	.084
b	0	.244	.209	.154	.114	.134	.067	.048	.030
	.05	.126	.102	.067	.045	.062	.025	.015	.011
	.10	.057	.044	.026	.025	.026	.008	.005	.003

For this data we could have a posterior probability of 0.741 of getting the best in the subset of suppliers 1, 2, 3, 5. Note that when using the $p_j(b)$ criterion supplier 5 is preferred to supplier 4 even though the sample proportion is in the reverse order. When using $E(\theta_j|\underline{x})$ this is not the case. This is reasonable in that the two criteria represent radically different goals for the experimenter. The $p_j(b)$ criterion should be used when a decision will be used for a long term and $E(\theta_j|\underline{x})$ should be used for the next lot. Further amplification of this point will be made later.

5. Remarks, discussion and conclusions

(i) Test of hypothesis

There may be some situations in which a decision maker is concerned in the first instance about testing the equality of the supplier's quality, i.e. test $H_0 : \theta_1 = \ldots = \theta_k$. Whereas we feel that this is not in general the ultimate goal of the experimenter, it is quite easy to incorporate this situation into the model by simply incorporating a prior probability γ that H_0 is true (i.e. $P(H_0 \text{ is true}) = P(\eta = 0) = \gamma$) and then computing the posterior probability of H_0 which is given by:

$$\gamma^* = \left[1 + \frac{1-\gamma}{\gamma}\frac{f(\underline{x})}{f(\underline{x}|0)}\right]^{-1}$$

where $f(\underline{x})$ and $f(\underline{x})|\beta,\eta)$ are given in (2.5) and

$$f(\underline{x}|0) = \int_0^1 f(\underline{x}|\beta,0)h_1(\beta)d\beta = \int_0^1 \prod_{i=1}^{k} \binom{n_i}{x_i} \beta^{\Sigma x_i}(1-\beta)^{N-\Sigma x_i} h_1(\beta)d\beta.$$

Then each p_j should be multiplied by $(1-\gamma^*)$ to obtain the posterior probability that θ_j is largest since p_j as given in (2.5) is conditional upon H_0 false, i.e. $\eta > 0$. One could simply compute the Bayes factor, $BF = f(\underline{x}|0)/f(\underline{x})$, as evidence for believing H_0. We point out however that the model of Deely and Zimmer (1987) seems more appropriate for testing the equality of supplier's quality.

(ii) Comparisons and possible extensions

It is clear that the HB model offers a much wider class of models than the naive Bayesian or the empirical Bayesian approaches which have been reported in the literature thus far. In the first instance, the HB model allows through the hyperpriors h_1 and h_2 the facility to use prior information about the suppliers as a group whereas the naive models have no place for such information. We believe that this prior information begins with an assumption of at least exchangeability, but more informative models are also possible as we have shown in the examples in Section 4. One could argue that some approximations of the p_j's or $E(\theta_i|\underline{x})$'s might be close enough and not require numerical integration. There has been some work in this direction (see Albert and Gupta (1985) and Leonard (1972)) but since the numerical integrations required herein are relatively easy such approximations would appear to be unnecessary. Secondly, we point out that only a very simple hierarchical model was used in this paper. It is clear that there is scope for richer models. For example, one could replace β in (2.1) with $y_{i1}\beta_1 + y_{i2}\beta_2$ where y_{i1}, y_{i2} are known "regressors" for $i = 1, \ldots, k$ and $\underline{\beta} = (\beta_1, \beta_2)$ is a vector of unknown "regression" coefficients with hyperprior $h_1(\underline{\beta})$. This model would incorporate various descriptions of changes in θ_i as well as the naive Bayesian model in which each θ_i is assumed independent with a known beta distribution possibly with different parameters. This latter case would be modeled by taking h_1 and h_2 as point distributions at (1, 1) and 1 respectively and then solving for y_{i1} and y_{i2} to obtain the given known beta parameters.

Another possible extension of the HB model would involve covering partial exchangeability particularly relevant when k is large. In this paper we have discussed analysis when k is small and have tacitly assumed all k binomial probabilities are exchangeable. It may be the case that, in a large group of suppliers, exchangeability is only tenable within sub-

groups and from subgroup to subgroup there may be exchangeability only in their means. Of course this fact may not be recognizable until after observing the data. The HB model should be enriched to allow the possibility of partial exchangeability being indicated by the data and then proceeding with the selection problem.

Finally it should be noted that the HB model has no difficulty with either small or variable sample sizes whereas naive empirical Bayes procedures require large sample sizes to imply their optimality properties. In addition these models cannot give practical answers to allocation of small samples amongst suppliers. In contrast the formulas for $p_j(b)$ or $E(\theta_j|\underline{x})$ developed herein can be used to generate a matrix of possibilities over a grid of varying small samples. The experimenter is then given tangible information by which a satisfactory design can be selected. There has been very little work done in this area. Recently, Yang (1988), has given sufficient conditions for $p_i(0) \leq p_j(0)$ as a function of x_i and x_j. He showed that if $x_j - x_i \geq \max(0, n_j - n_i)$ then $p_j(0) \geq p_i(0)$. Although this condition is useful, it does not completely partition the (x_i, x_j) space and in fact when $n_j - n_i$ is large there are many possibilities for x_i and x_j which do not satisfy Yang's condition. In particular the region where $(x_i/n_i) = (x_j/n_j)$ (or nearly so) does not in general satisfy this condition. Our numerical results seem to indicate that over this region the smaller sample size gives the larger $p_j(0)$; but this remains to be demonstrated completely.

(iii) Differences in selection criteria

It has been proposed in this paper that either the $p_j(b)$'s or the $E(\theta_j|\underline{x})$'s be used for selection purposes. Which to use will depend on the requirements of the practical situation. If a decision is to be made, say contracting with the selected suppliers for delivery of items over a period of time, then $p_j(b)$ should be used for either selecting the best or selecting

the smallest subset for which the posterior probability that the largest (by amount b) θ_j is in the selected subset is at least P^*, i.e. the PP^* rule. If, however, a decision for the short term is to be made, say which machine to use for the next n items, then $E(\theta_j|\underline{x})$ is more appropriate. To select a subset using this criterion, the requirement could be either to insure that the expected number of successes is at least N^* (i.e. take $c = N^*/n$ in Section 2 (ii)) or to maximize the expected number of successes from a fixed number r of the k suppliers, $r < k$.

It should be noted that if a decision theoretic approach is taken for the subset selection problem, there is no known loss function which gives as the Bayes procedure the PP^* rule. Gupta and Yang (1985) give general conditions which must be satisfied by the loss function in order that the PP^* is Bayes amongst the restricted class of rules satisfying the P^* condition. If the decision problem is formulated as selecting a subset of fixed size then the procedure discussed above based on $E(\theta_j|\underline{x})$ is Bayes with respect to, say

$$L(S_r, \underline{\theta}) = k\theta_{[k]} - \sum_{i \in S_r} \theta_i$$

where S_r ranges over subsets of size r. However the procedure which insures the expected number of successes is at least N^* has not yet been shown to be a Bayes procedure in the decision theoretic sense.

It should also be noted that the two criteria can lead to different subsets being selected, as shown in Section 4(iii). This is even true when only a single supplier is to be selected. This is not surprising since the two criteria clearly have different objectives as discussed earlier. Furthermore the $p_j(b)$ calculation depends on the variance as well as the mean so when sample sizes are quite different but proportions similar, it is to be expected that the largest $p_j(b)$ does not correspond to the largest $E(\theta_j|\underline{x})$.

(iv) Conclusions

We have tried to show why the HB model is helpful to the practical problem of selecting the best amongst k binomial populations. The salient features of this approach are:

(i) the ability to deal easily with variable and small sample sizes;

(ii) the incorporation in the model of prior information concerning the suppliers as a group;

(iii) the ease of computation of the selection criteria;

(iv) the dependence of the optimality qualities upon differences in the observations as opposed to differences in the unobserved parameter space.

Further work remains to be done to make these techniques available to the experimenter, but we hope we have made some progress in that direction.

References

Albert, J.H. (1984), "Empirical Bayes estimation of a set of binomial probabilities", *Journal of Statistical Computation and Simulation*, 20, 129–144.

Albert, J.A. and Gupta, A.K. (1985), "Bayesian methods for binomial data with applications to a nonresponse problem", *Journal of the American Statistical Association*, 80, 167-174.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

Berger, J.O. and Deely, J.J. (1988), "A Bayesian approach to ranking and selection of related means with alternatives to AOV methodology", *Journal of the American Statistical Association* (in press).

Bratcher, T.L. and Bland, R.P. (1975), "On comparing binomial probabilities from a Bayesian viewpoint", *Communications in Statistics*, 4, 975–985.

Brooks, R.J. (1987), "Optimal allocation for Bayesian inference about an odds ratio", *Biometrika*, 74, 196–199.

Brittain, E. and Schlesselman, J.J. (1982), "Optimal allocation for the comparison of proportions", *Biometrics*, 38, 1003–1009.

Deely, J.J. (1965), Multiple decision procedures from an empirical Bayes approach", Ph.D. (Mimeo. Series No. 45), Dept. of Statistics, Purdue University, West Lafayette, Indiana.

Deely, J.J. and Lindley, D.V. (1981), "Bayes empirical Bayes", *Journal of the American Statistical Association*, 76, 833–841.

Deely, J.J. and Zimmer, W.J. (1988), "Choosing a quality supplier – a Bayesian approach",

Bayesian Statistics III, (Eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith), In press.

Eaves, D.M. (1980), "On exchangeable priors in lot acceptance", *J. Roy. Statist. Soc.* B, 42, 88–93.

Gibbons, J.D., Olkin, I. and Sobel, M. (1977), *Selecting and Ordering Populations*, New York: Wiley.

Gupta, S.S., Huang, D.-Y. and Huang, W.-T. (1976), "On ranking and selection procedures and tests of homogeneity for binomial populations", *Essays in Probability and Statistics* (Eds. S. Ikeda, et al.), Shinko Tsusho Co. Ltd., Tokyo, Japan, pp. 501–533.

Gupta, S.S. and Liang, T. (1987), "Empirical rules for selecting the best binomial population", *Statistical Decision Theory and Related Topics – IV* (Eds. S.S. Gupta and J.O. Berger), Vol. 1, Springer-Verlag, New York, pp. 213–224.

Gupta, S.S. and Sobel, M. (1960), "Selecting a subset containing the best of several binomial parameters", *Contributions to Probability and Statistics*, (Eds. I. Olkin et al.), Stanford University Press, Stanford, California, pp. 224–248.

Gupta, S.S. and Panchapakesan, S. (1979), *Multiple Decision Procedures*, New York: Wiley.

Gupta, S.S. and Yang, H.M. (1985), "Bayes-P^* subset selection procedures for the best population", *Journal of Statistical Planning and Inference*, 12, 213–233.

Leonard, T. (1972), "Bayesian methods for binomial data", *Biometrika*, 59, 581–589.

Morris, C.N. (1983), "Parametric empirical Bayes inference: theory and applications", *Journal of the American Statistical Association*, 78, 47–65.

Sobel, M. and Huyett, M.J. (1957), "Selecting the best one of several binomial popula-

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report #88-21C	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Purdue University / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) Department of Statistics West Lafayette, IN 47907	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-88-K-0170, DMS-8606964 DMS-8702620
8c. ADDRESS (City, State, and ZIP Code) Arlington, VA 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Hierarchical Bayesian Selection Porcedures for the Best Binomial Population (unclassified)

12. PERSONAL AUTHOR(S)
John J. Deely and Shanti S. Gupta

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM ____ TO ____	May, 1988	22

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	binomial, Bayes, hierarchical, selection procedures, exchangeable, hyperpriors, empirical Bayes

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In this paper a hierarchical Bayesian (HB) model is adopted to derive selection procedures for selecting the best of k binomial parameters, say theprobability of success corresponding to k different suppliers. This model facilitates the use of prior information in the analysis for both small and large sample sizes. In addition to computing posterior probabilities that the i^{th} supplier is best, this paper presents expressions for deciding how much better a given supplier is relative to the others. Prior information is assumed to begin with exchangeability and can be more informative if the experimenter has other knowledge about the suppliers as a group. A numerical example is given and the paper concludes with remarks about future work.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL John J. Deely and Shanti S. Gupta	22b. TELEPHONE (Include Area Code) (317) 494-6030 / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

tions", *Bell System Technical Journal*, 36, 537–576.

Yang, H.M. (1987), *"On selecting the treatment with largest probability of survival"*, submitted for publication.